

Tanvir Ahmed Khan

Research Interests

- Areas Compilers, Operating Systems, Computer Architecture
Topics Profile-guided Optimization, Locality Optimization, Prefetching

Education

- 2017-present **Ph.D.**, *University of Michigan*, Ann Arbor, Michigan, USA.
○ Computer Science and Engineering ○ CGPA: 3.775/4.00
○ Thesis: Profile-guided Locality Optimization for Data Center Applications
○ Advisor: Prof. Baris Kasikci
- 2014-2017 **M.Sc.**, *Bangladesh University of Engineering and Technology*, Dhaka, Bangladesh.
○ Computer Science and Engineering ○ CGPA: 3.75/4.00
○ Thesis: Overcoming Throughput Degradation in Multi-Radio Cognitive Radio Networks
○ Advisor: Dr. A. B. M. Alim Al Islam
- 2009-2014 **B.Sc.**, *Bangladesh University of Engineering and Technology*, Dhaka, Bangladesh.
○ Computer Science and Engineering ○ CGPA: 3.97/4.00
○ Thesis: Too Many Cooks Spoil the Broth: Augmenting Spectral Harvesting with Multiple Radios Can Make the Performance Worse!
○ Advisor: Dr. A. B. M. Alim Al Islam ○ Class Rank: 1/153

Selected Publications

- OSDI'21 **Tanvir Ahmed Khan**, Ian Neal, Gilles Pokam, Barzan Mozafari, and Baris Kasikci, DMon: Efficient Detection and Correction of Data Locality Problems using Selective Profiling. in proceedings of the 15th USENIX Symposium on Operating Systems Design and Implementation (OSDI), USENIX Association, 2021.
- ISCA'21 **Tanvir Ahmed Khan**, Dexin Zhang, Akshitha Sriraman, Joseph Devietti, Gilles Pokam, Heiner Litz, and Baris Kasikci, Ripple: Profile-Guided Instruction Cache Replacement for Data Center Applications. in proceedings of the 48th International Symposium on Computer Architecture (ISCA), 2021.
- FAST'21 Ian Neal, Gefei Zuo, Eric Shiple, **Tanvir Ahmed Khan**, Youngjin Kwon, Simon Peter, and Baris Kasikci, Rethinking File Mapping for Persistent Memory. in proceedings of the 19th USENIX Conference on File and Storage Technologies (FAST), USENIX Association, 2021.
- MICRO'20 **Tanvir Ahmed Khan**, Akshitha Sriraman, Joseph Devietti, Gilles Pokam, Heiner Litz, and Baris Kasikci, I-SPY: Context-Driven Conditional Instruction Prefetching with Coalescing. in proceedings of the 53rd IEEE/ACM International Symposium on Microarchitecture (MICRO), IEEE, 2020.
- IISWC'20 Yuhan Chen, Jingyuan Zhu, **Tanvir Ahmed Khan**, and Baris Kasikci, CPU Microarchitectural Performance Characterization of Cloud Video Transcoding. in proceedings of the IEEE International Symposium on Workload Characterization (IISWC), IEEE, 2020.

- PLDI'19 **Tanvir Ahmed Khan**, Yifan Zhao, Gilles Pokam, Barzan Mozafari, and Baris Kasikci, Huron: Hybrid False Detection and Repair. in proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI), ACM, Phoenix, Arizona, United States, pages 473-488, 2019.
- MobiSys'16 **Tanvir Ahmed Khan** and A. B. M. Alim Al Islam, Poster: Overcoming Throughput Degradation in Multi-Radio Cognitive Radio Networks. in proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys) Companion, 2016, ACM, Singapore, Singapore, pages 41-41, 2016.
- WiMob'15 **Tanvir Ahmed Khan**, Chowdhury Sayeed Hyder, and A. B. M. Alim Al Islam, Towards exploiting a synergy between cognitive and multi-radio networking. in proceedings of the 11th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 2015, IEEE, Abu Dhabi, UAE, pages 370-377, 2015.

Awards and Honors

- 2020 Facebook Fellowship Finalist
- 2019 PLMW@PLDI 2019 Student Scholarship covering conference registration, accommodation, and travel.
- 2017-2018 Rollin M. Gerstacker Foundation Fellowships, University of Michigan.
- 2018 OSDI 2018 Student Grant covering conference registration, accommodation, and travel.
- 2016 MobiSys 2016 Student Scholarship covering conference registration, accommodation, and travel.
- 2014 Crest of Honor, Highest CGPA in the department, presented by BUET alumni association.
- 2009-2014 University Merit Scholarship, Bangladesh University of Engr and Tech.
- 2009-2014 Dean's List Scholarship, Bangladesh University of Engr and Tech.

Employment

- 2017-present **Research Assistant**, *University of Michigan*, Ann Arbor, Michigan, USA.
- Advisor: Baris Kasikci
 - Electrical Engineering and Computer Science Department.
- Summer 2020 **Software Engineer Intern**, *Facebook*, Menlo Park, California, USA.
- Mentor: Maksim Panchenko
 - Binary Optimization and Layout Tool (BOLT) Team
- Summer 2019 **Research Intern**, *Microsoft*, Redmond, Washington, USA.
- Mentor: Gagan Gupta and Rathijit Sen
 - Azure Hardware Research Group
- Winter 2019 **Graduate Student Instructor**, *University of Michigan*, Ann Arbor, Michigan, USA.
- Electrical Engineering and Computer Science Department.
 - Primary Instructor: Thomas Wensch
 - Course: Parallel Computer Architecture.
- 2014-2017 **Lecturer**, *Bangladesh University of Engineering and Technology*, Dhaka, Bangladesh.
- Department of Computer Science and Engineering.
 - Courses Taught: Operating Systems, Computer Architecture, Compilers.

Talks

- 2020 **I-SPY: Context-Driven Conditional Instruction Prefetching with Coalescing**.
- IEEE/ACM International Symposium on Microarchitecture (MICRO), October 2020

- 2019 **Huron: Hybrid False Sharing Detection and Repair.**
 - Microsoft C++ Team, August 2019
 - ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI), June 2019
 - Azure Hardware Research Group, May 2019
- 2018 **Overcoming Throughput Degradation in Multi-Radio Cognitive Radio Networks.**
 - Intel Labs Wireless Networking Research Group, May 2018

Selected Research Projects

- 2020-Present **Profile-guided Binary Layout Optimization for Linux Kernel.**

Many data center applications suffer from frequent instruction cache misses costing millions of dollars. While existing profile-guided optimization mechanisms can improve the instruction locality of user-mode instructions for these applications, no such technique exists for kernel-mode instructions. In this work, we first characterize the high percentage of instruction cache misses coming from kernel mode instructions. Then, we enhance the state-of-the-art profile-guided optimization tool, BOLT to rewrite the Linux Kernel binary.
- 2019-Present **Context-Driven Conditional Instruction Prefetching with Coalescing for Data Center Applications.**

Modern data center applications exhibit large instruction footprints, leading to frequent instruction cache misses increasing cost and degrading data center performance and energy efficiency. To overcome this performance degradation via prefetching, we first investigate the challenges of effective instruction prefetching. We then use insights derived from our investigation to develop I-SPY, a novel profile-guided prefetching technique. Specifically, I-SPY proposes context-driven conditional prefetching and prefetch coalescing. For nine data center applications, I-SPY provides an average of 15.5% speedup and 95.9% reduction in instruction cache misses outperforming the state-of-the-art prefetching technique by 22.5%.
- 2019-Present **Performance Characteristics of Managed Workloads.**

Many modern applications are written in managed code to achieve type safety, portability, security, and automatic memory management. While these features provide necessary abstractions for quick and safe software development, computer architectural aspects of these features are not well studied in the literature. In this study, we investigate the characteristics of these managed workloads. We identify the key bottlenecks that significantly hurt the performance of these workloads on modern hardware. We also correlate these bottlenecks on different aspects of managed runtime including common shared libraries, garbage collection, and just in time code generation.
- 2018-Present **Continuous Locality Optimization.**

Locality of reference is a key factor affecting a program's performance. However, very few programs exhibit good data locality during execution. In this project, we propose Erie, the first methodology that can continuously monitor executions in production for locality optimizations, with negligible overhead and without requiring special hardware. Using a hybrid approach, Erie efficiently identifies locality problems in production, and addresses them in house. Erie can reason about how its optimizations affect unobserved executions, as well as their interaction with underlying hardware mechanisms. Our evaluation shows that Erie's overhead is less than 1.36% on average, and its optimizations provide up to 53.14% (16.83% on average) speedup, compared to a baseline that uses the highest level of compiler optimization. Erie provides up to 17.48% (6.64% on average) speedup on PostgreSQL, one of the most popular open-source data-base management systems.
- 2017-2019 **Automatic False Sharing Elimination.**

Parallel programs on shared memory multiprocessors suffer from a slowdown of an order of magnitude due to false sharing. False sharing occurs when multiple threads demand exclusive access to a cache line even though their data locate on non-overlapping portions of the cache line. In this project, we proposed Huron, a hybrid in-house/in-production false sharing detection and repair system. Huron detects and repairs as much false sharing as it can in-house, and relies on its lightweight in-production mechanism for remaining cases. The key idea behind Huron's in-house false sharing repair is to group together data that is accessed by the same set of threads, to shift falsely-shared data to different cache lines. Huron's in-house repair technique can generalize to previously-unobserved inputs. Our evaluation shows that Huron can detect more false sharing bugs than all state-of-the-art techniques, and with a lower overhead. Huron improves runtime performance by 3.82× on average (up to 11×), which is 2.11-2.27× better than the state of the art.

2013-2017 **Performance Analysis for Multi-Radio Cognitive Radio Networks.**

In this project, we analyzed several performance metrics (throughput, delay, packet-loss, etc.) in multi-radio cognitive radio networks. Consequently, we proposed a feedback-based multi-radio exploitation approach to overcome throughput degradation in multi-radio cognitive radio networks.

Selected Services

- 2021 External Review Committee Member for ASPLOS'21, Shadow Program Committee Member for EuroSys'21
- 2020 Artifact Evaluation Committee Member for PLDI'20, MLSys'20, and ASPLOS'20
- 2019 Artifact Evaluation Committee Member for SOSP'19, Student Volunteer for PLDI'19

Grants

- 2020 Taming the Instruction Bottleneck in Modern Datacenter Applications
 - Principal Investigators: Baris Kasikci and Joseph Devietti
 - Organization: NSF/Intel Partnership on Foundational Microarchitecture Research (FoMR)
- 2020 Proxy-Web: A Proxy App Suite for Production Web Services
 - Principal Investigators: Baris Kasikci, Timothy Rogers, and David Brooks
 - Organization: Semiconductor Research Corporation (SRC)

Selected Research Mentoring

- 2020 Dexin Zhang, undergrad, co-author of the ISCA'21 paper
- 2019-Present Nathan Brown, undergrad, first place in ACM undergraduate student research competition, CGO'20
- 2020 Yuhan Chen, PhD student, results published in IISWC'20
- 2018-19 Yifan Zhao, undergrad (now a PhD student, UIUC), co-author of the PLDI'19 paper

Technical Skills

- Languages C, C++, Java, x86 Assembly, PL/SQL, Tcl, Python, HTML, CSS, JavaScript, PHP, Shell script, Prolog.
- DB Systems PostgreSQL, MySQL, SQLite.
- Miscellaneous NS-2, NS-3, LLVM, Intel Pin, SimPy, ZSim, Awk, Flex/Bison, OpenGL, Linux perf, Intel PMU tools, Intel Tools PT, nvprof, pgprof, valgrind, Spring, Android, Bootstrap, Arduino, TensorFlow, \LaTeX .